Banff International Research Station
for Mathematical Innovation and Discovery

Sparse Random Structures: Analysis and Computation
January 24 – 29, 2010

MEALS

*Breakfast (Buffet): 7:00–9:30 am, Sally Borden Building, Monday–Friday
*Lunch (Buffet): 11:30 am–1:30 pm, Sally Borden Building, Monday–Friday
*Dinner (Buffet): 5:30–7:30 pm, Sally Borden Building, Sunday–Thursday
Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall
***Please remember to scan your meal card at the host/hostess station in the dining room for each meal.**

MEETING ROOMS

All lectures will be held in Max Bell 159 (Max Bell Building accessible by walkway on 2nd floor of Corbett Hall). LCD projector, overhead projectors and blackboards are available for presentations. Please note that the meeting space designated for BIRS is the lower level of Max Bell, Rooms 155–159. Please respect that all other space has been contracted to other Banff Centre guests, including any Food and Beverage in those areas.

SCHEDULE

Sunday

16:00	Check-in begins (Front Desk - Professional Development Centre - open 24 hours) Lecture rooms available after 16:00 (if desired)
17:30–19:30	Buffet Dinner, Sally Borden Building

Monday

7:00–8:45	Breakfast
8:45–9:00	Introduction and Welcome to BIRS by BIRS Station Manager, Max Bell 159
9:00–9:30	John Gilbert
	Challenges in Combinatorial Scientific Computing
9:30–10:00	Al Hero
	Large scale correlation screening
10:00–10:30	Tim Davis
	The University of Florida Sparse Matrix Collection
10:30–11:00	Coffee Break, 2nd floor lounge, Corbett Hall
11:00–11:30	Charles Bordenave
	Spectrum of non-hermitian heavy tailed random matrices
11:30–12:00	Brendan D. McKay
	Combinatorial estimates by the switching method
12:00–13:00	Lunch
13:00–14:00	(Optional) Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall
14:00 –14:30	Jure Lescovec
	Modeling graphs with node attributes
14:30 – 15:00	Arnab Sen
	Spectra of Large Random Trees
15:00 – 15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30 – 16:00	Tim Rogers
	Sparse and Non-Hermitian Random Matrix Theory
16:00 – 16:30	Philippe Biane
	Brown measure for some sums of free operators
16:30 – 17:00	Jonathan Novak
	Restricted permutation matrices and Mehta's integral
17:00 – 17:30	Balint Virag
	Spiked matrices, the stochastic Airy operator and the Painleve equation
17:30–19:30	Dinner

Tuesday

7:00–8:45	Breakfast
9:00–9:30	Ian Blake
	Coding for erasures and random matrices
9:30–10:00	Mark Rudelson
	Random conjunctions matrices
10:00– 10:30	Michael W. Mahoney
	Learning on large informatics graphs
10:30–11:00	Coffee Break, 2nd floor lounge, Corbett Hall
11:00 – 11:30	Tamara G. Kolda
	Scalable Tensor Factorizations with Incomplete Data
11:30 – 12:00	Rachel Ward
	Sparse Legendre expansions via l1 minimization
12:00–12:15	Group Photo outside Corbett Hall
12:15–13:30	Lunch
14:00– 14:30	Stephen A. Vavasis
	Convex relaxation for the clique, biclique and clustering problem
14:30– 15:00	Hristo Djidjev
	Identifying the community structure of large networks
15:00 – 15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30 – 16:00	Patrick Hayden
	Applications and desiderata for random matrices in quantum information
16:00 – 16:30	Matthew Harding
	Identification of Communities in High Dimensional Stochastic Networks
16:30 – 17:00	Victor Preciado
	How Relevant are Local Measurement in Complex Network Modeling?
17:30–19:30	Dinner

Wednesday

7:00–8:45	Breakfast
9:00–9:30	Per-Gunnar Martinsson
	Enabling very large-scale computations via randomization
9:30–10:00	Raj Rao Nadakuditi
	Eigenvalue and eigenvector phase transitions due to perturbation or randomization
10:00– 10:30	Yuliy Baryshnikov
	On the statistical topology of hard disks in a box
10:30–11:00	Coffee Break, 2nd floor lounge, Corbett Hall
11:00 – 11:30	James Mingo
	Sharp Bounds for Sums Associated to Graphs of Matrices
11:30 – 12:00	Alan Edelman
	Shape Theory Meets Random Matrix Theory with Geometry, History and Applications
12:15– 13:30	Lunch
13:30 – 17:30	Free afternoon - group excursion
17:30 – 19:30	Dinner

Thursday

7:00–9:00	Breakfast
9:00–9:30	Ioana Dumitriu *On the spectrum and eigenvectors of large regular random graphs*
9:30–10:00	Patrick Perry *Cross-Validation for Principal Components Analysis*
10:00– 10:30	Yaniv Plan *Tight Oracle Bounds for Low-rank Matrix Recovery from a Minimal Number of Random Measurements*
10:30–11:00	Coffee Break, 2nd floor lounge, Corbett Hall
11:00 – 11:30	Alexander Litvak *RIP of matrices with independent columns and neighborly polytopes*
11:30 – 12:00	Po-Ru Loh *A numerical linear algebra view of the Tao-Vu smallest singular value limit*
12:00–13:30	Lunch
14:00– 14:30	Brian Rider *Small deviations for beta ensembles*
14:30– 15:00	Stephen Curran *Probabilistic aspects of easy quantum groups*
15:00 – 15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30 – 16:00	Wlodek Bryc *Meixner matrix ensembles*
16:00 – 16:30	Eugene Kritchevski *The scaling limit of the critical one-dimensional random Schrödinger operator*
16:30 – 16:45	Closing remarks
17:30–19:30	Dinner

Friday

7:00–9:00	Breakfast
9:00	Informal Discussions
	Coffee Break, 2nd floor lounge, Corbett Hall - to START no earlier than 10 am
11:30–13:30	Lunch

Checkout by 12 noon.

** 5-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon (unless later departure has been previously requested and explicitly approved by BIRS Station Director). **

Banff International Research Station

for Mathematical Innovation and Discovery

Sparse random structures: Analysis and Computation
January 24 – 29, 2010

ABSTRACTS
(in alphabetic order by speaker surname)

Speaker: **Yuliv Baryshnikov** (Bell Labs)
Title: *On the statistical topology of hard disks in a box*
Abstract: Elastically colliding hard balls (or disks) is the archetypal dynamical system of statistical mechanics. While a lot of attention was paid to the question of ergodicity (mixing etc) of the corresponding flow, the far more basic question of the topology of the underlying configuration space (say, is it connected? not always!) was mostly ignored. I will address the question of the asymptotic Betti numbers of these configuration spaces.

Speaker: **Philippe Biane** (University Marne le Valle)
Title: *Brown measure for some sums of free operators*
Abstract: The Brown measure of an operator is an adequate substitute, in an operator algebraic context, for the empirical spectral distribution of a non self-adjoint element. We give some examples of computation of such Brown measures in free probability, which should give explicit formulas for the limit of the empirical spectral distribution of some natural models of random matrices. This is joint work with F. Lehner.

Speaker: **Ian Blake** (University of Toronto)
Title: *Coding for erasures and random matrices*
Abstract: Much research on coding for the erasure channel over the past decade has focused on random coding techniques resulting in the so-called raptor or online codes. These codes reduce encoding and decoding complexity from cubic (in number of information packets) to linear and find application to the problem of internet packet loss channels. This work reviews the key elements of these coding techniques and considers them from the perspective of the rank properties of random matrices. Several conjectures are made on the behavior of random matrices over finite fields where the matrix elements are chosen iid with the probability of a nonzero element p.

Speaker: **Charles Bordenave** (University Toulouse)
Title: *Spectrum of non-hermitian heavy tailed random matrices*
Abstract: Consider a n x n matrix with independent and identically distributed entries. If the variance is finite, the circular law Theorem asserts that the empirical spectral distribution converges to the uniform distribution on the unit complex disc. This theorem was recently proved by Tao and Vu after important breakthroughs of Girko, Bai and others. In this talk, we will consider the infinite variance case. We will state the convergence of the empirical spectral measure and give some properties of the limiting distribution. This is a joint work with Djalil Chafai and Pietro Caputo.

Speaker: **Wlodek Bryc** (University of Cincinnati)
Title: *Meixner matrix ensembles*
The talk is based on joint work (in progress) with Gerard Letac. We show that the Laplace transform of

an n by n matrix ensemble that fits Anshelevich's regression postulates for "Meixner laws on matrices" satisfies a system of PDEs which is explicitly solvable for n=2. By constructing the corresponding 2 by 2 Meixner matrix ensembles, we establish that for appropriate values of parameters these solution are indeed the Laplace transforms. This is joint work with Gerard Letac.

Speaker: **Stephen Curran** (University of California, Berkeley)
Title: *Probabilistic aspects of easy quantum groups*
Abstract: The class of "easy" quantum groups was introduced by Banica and Speicher to provide a framework for studying certain common probabilistic features of S_n, O_n and their "free versions", due to Wang, which are compact quantum groups in the sense of Woronowicz. In this talk I will survey some recent results around this family of quantum groups, including de Finetti theorems and extensions of some results of Diaconis-Shahshahani. This is joint work with Teodor Banica and Roland Speicher.

Speaker: **Tim Davis** (University of Florida)
Title: *The University of Florida Sparse Matrix Collection*
Abstract: The University of Florida Sparse Matrix Collection is a large, widely available, and actively growing set of sparse matrices that arise in real applications. Its matrices cover a wide spectrum of domains, include those arising from problems with underlying 2D or 3D geometry (such as structural engineering, computational fluid dynamics, model reduction, electromagnetics, semiconductor devices, thermodynamics, materials, acoustics, computer graphics/vision, robotics/kinematics, and other discretizations) and those that typically do not have such geometry (such as optimization, circuit simulation, economic and financial modeling, theoretical and quantum chemistry, chemical process simulation, mathematics and statistics, power networks, and other networks and graphs). The collection is widely used by the sparse matrix algorithms community for the development and performance evaluation of sparse matrix algorithms. The collection includes software for accessing and managing the collection, from MATLAB, Fortran, and C. This is joint work with Yifan Hu (AT & T)

Speaker: **Hristo Djidjev** (Los Alamos National Laboratories)
Title: *Identifying the community structure of large networks*
Abstract: One way to analyze and understand the structure and the functioning of large networks is to divide their nodes into communities/clusters – maximal groups of nodes with denser in-cluster links and fewer links connecting nodes from different clusters. One approach to finding the communities of a given network is to look for a partition that maximizes its modularity, which is proportional to the difference between the number of the existing links between nodes from the same part and the expected number of such links in a random network. Most often used random network models include the Erdos-Renyi G(n,p) model, where each edge appears with the same probability p, and the Aiello-Chung-Lu model, where the probability depends on the degrees of its endpoints. In this talk we will show that the problem of finding a partition maximizing the modularity of a given network N can be reduced to solving a number of minimum weighted cut problems on a complete graph N', where N' has the same vertices as N and appropriately defined edge weights. We will then demonstrate that the resulting minimum cut problem can be efficiently solved using multi-level graph partitioning methods and that our algorithm finds clusterings of comparable or better quality and much faster than the existing community detection algorithms.

Speaker: **Ioana Dumitriu** (University of Washington)
Title: *On the spectrum and eigenvectors of large regular random graphs*
Abstract: We examine the asymptotic behavior of the spectrum and eigenvectors of adjacency matrices of d-regular random n-graphs in the regimes where $d = n^{\epsilon_n}$ and $d = (\log n)^\gamma$ with $\gamma < 1$. We prove convergence of the empirical spectral distribution to the semicircle law for the first and examine the rate of convergence on small intervals in the case of the second. We also examine the eigenvectors, and conclude that they obey a couple of properties (delocalization, lack of bias) that support the conjecture that their distributions are close to the uniform distribution on the sphere.

Speaker: **Alan Edelman** (MIT)
Title: *Shape Theory Meets Random Matrix Theory with Geometry, History and Applications*
Abstract: Shapes come from matrices. We would be pleased if this talk helps you to see the familiar triangle in a new light. We would be even more pleased if we could help encourage applications perhaps with tools from random matrixtheory. We began with an idle question and a note passed in lecture. (We do not condone our behavior!) The question was random: what is the probability that a triangle is acute? The note contained an integral over a region in $\mathbb{R}^6$.

Are most random triangles acute or obtuse? Later we learned that Lewis Carroll (as Charles Dodgson) asked the same question. It has been answered multiple times, and in every case obtuse triangles are the winners – if our mental image of a typical triangle is acute, we are wrong. The margin of victory, of course, depends on the meaning of the word random. Probably a triangle taken randomly from a high school geometry book would indeed be acute. We suppose that humans probably think of acute triangles, indeed equilateral or nearly equilateral triangles in our mental representations of a generic triangle.

The most developed piece of the subject is humbly known as "Shape Theory." It was the last interest of the first professor of mathematical statistics at Cambridge University, David Kendall [Kendall99]. We rediscovered on our own what theshape theorists knew, that triangles are naturally mapped onto points of the sphere. It was a thrill to discover both the result and the history, and we hope here to contribute something new as well.

We will add to the theory a purely geometrical derivation of the triangle to sphere link, delve into the linear algebra point of view, and opportunities to rejuvenate the theory with new applications and connections to random matrix theory. This is joint work with Gilbert Strang at MIT

Speaker: **John R. Gilbert** (University of California, Santa Barbara)
Title: *Challenges in Combinatorial Scientific Computing*
Abstract: Computation on large combinatorial structures – graphs, strings, partial orders, etc. – has become fundamental in many areas of data analysis and scientific modeling. The field of high-performance combinatorial computing, however, is in its infancy. By way of contrast, in numerical supercomputing we possess standard algorithmic primitives, high-performance software libraries, powerful rapid-prototyping tools, and a deep understanding of effective mappings of problems to high-performance computer architectures. This talk will describe several challenges for the field of combinatorial scientific computing in algorithms, tools, architectures, and mathematics. I will draw examples from several applications, and I will highlight our group's work on high-performance implementation of algebraic primitives for computation on large graphs.

Speaker: **Matthew Harding** (Stanford University)
Title: *Identification of Communities in High Dimensional Stochastic Networks*
Abstract: In this paper, we present a framework for the non parametric identification of clusters / communities in high dimensional stochastic networks. We assume that the adjacency matrix at any instant of time is a linear combination of the true adjacency matrix and a noise matrix. In addition, we assume that the noise matrix is a Wigner matrix (Bai et al., 1999) . In this paper, we employ a random matrix theory approach for the identification problem. To this end, we will use the Stieltjes Transform to obtain closed form solutions of the moments of the eigenvalue distribution function of the noise matrix. We then develop an algorithm using these theoretical moments and a general- ized method of moments (GMM) methodology to uncover the underlying deterministic structure of linking within various communities. In addition, we will use results from matrix perturbation theory to help in the assessing the robustness of the identification algorithm. Next, we use the developed algorithm on simulated datasets as well as real life data. This is joint work with K. Krishnan Nair.

Speaker: **Patrick Hayden** (McGill University)
Title: *Applications and desiderata for random matrices in quantum information*
Abstract: As in "classical" computer science, the probabilistic method and randomized constructions

are ubiquitous in quantum information. Because quantum mechanics is noncommutative, the random objects of study are invariably matrices. Quantum information theory therefore provides a rich source of random matrix problems. Applications range from the best known codes for sending quantum data through noisy media to subroutines in quantum algorithms and new encryption procedures. The first step in developing these ideas usually involves demonstrating that a very large random unitary transformation selected according to the Haar measure accomplishes the appropriate task. The next, and often much harder step, is to find an efficiently constructible replacement for the unwieldy Haar-measure unitaries. I'll illustrate some of these applications, describe the progress that has been made on efficient constructions and pose some open problems.

Speaker: **Al Hero** (University of Michigan)
Title: *Large scale correlation screening*
Abstract: The problem of screening a large number of variables for pairwise correlations is a classical statistical problem that has been studied for several decades. This problem arises in many disciplines including gene expression analysis, finance, and security where the number of variables can range from a few hundred to hundreds of thousands. We distinguish between three correlation screening applications: screening auto-correlations in a single treatment, screening cross-correlations in multiple treatments, and screening persistent auto-correlations in multiple treatments. In this talk we will discuss fundamental limits of large scale correlation screening for these applications.

Speaker: **Tamara G. Kolda** (Sandia National Laboratories)
Title: *Scalable Tensor Factorizations with Incomplete Data*
Abstract: The problem of incomplete data—i.e., data with missing or unknown values—in multi-way arrays is ubiquitous in biomedical signal processing, network traffic analysis, bibliometrics, social network analysis, chemometrics, computer vision, communication networks, etc. We consider the problem of how to factorize data sets with missing values with the goal of capturing the underlying latent structure of the data and possibly reconstructing missing values (i.e., tensor completions). We focus on one of the most well-known tensor factorizations that captures multi-linear structure, CANDECOMP/PARAFAC (CP). In the presence of missing data, CP can be formulated as a weighted least squares problem that models only the known entries. We develop an algorithm called CP-WOPT (CP Weighted OPTimization) that uses a first-order optimization approach to solve the weighted least squares problem. Based on extensive numerical experiments, our algorithm is shown to successfully factorize tensors with noise and up to 99% missing data. A unique aspect of our approach is that it scales to sparse large-scale data, e.g., 1000 x 1000 x 1000 with one million known entries (0.1% dense). To show the real-world usefulness of CP-WOPT, we illustrate its applicability on a novel EEG (electroencephalogram) application where missing data is frequently encountered due to disconnections of electrodes and also on the problem of modeling network traffic where data may be absent due to collection errors. This is joint work with Evrim Acar and Daniel M. Dunlavy.

Speaker: **Eugene Kritchevski** (University of Toronto)
Title: *The scaling limit of the critical one-dimensional random Schrödinger operator*
Abstract: We study the one dimensional discrete random Schroödinger operator

$$(H_N\psi)_n = \psi_{n-1} + \psi_{n+1} + v_n\psi_n,$$

$\psi_0 = \psi_{N+1} = 0$, in the scaling limit $\mathrm{Var}(v_n) = 1/N$. We show that, in the bulk of spectrum, the eigenfunctions are delocalized and that there is a very strong repulsion of eigenvalues. The analysis is based on a stochastic differential equation for the evolution of products of transfer matrices. The talk is based on a joint work with Benedek Valko and Balint Virag.

Speaker: **Jure Lescovec** (Stanford University)
Title: *Modeling graphs with node attributes*

Abstract: Real-world networks from many domains exhibit a large number of structural similarities, such as the form of the distributions of node degrees, the spectrum of the adjacency matrix, and the degree of interconnectedness. While many generative models attempt to capture various aspects of network structure they usually neglect the heterogeneity in the population of nodes. In this talk I will introduce and investigate a binomial attribute graphs, a new model inspired by the Kronecker graph construction. Surprisingly, despite discarding a large amount of structure imposed by the Kronecker graph construction, the new model performs at least as well, and is much more easily modified, as it follows standard probabilistic semantics. Binomial attribute graphs are too simplistic to serve as general-purpose network models, but they provide new insight into why the Kronecker graph construction works and suggest directions for improvement.

Speaker: **Alexander Litvak** (University of Alberta)
Title: *RIP of matrices with independent columns and neighborly polytopes*
Abstract: We consider compressed sensing matrices and neighborliness of centrally symmetric convex polytopes generated by vectors $\pm X_1, \ldots, \pm X_N \in \mathbb{R}^n$, $(N \geq n)$. We introduce a class of random sampling matrices and show that with overwhelming probability they are valid for the exact reconstruction process of m-sparse vectors via ℓ_1 minimization. In particular, matrices with i.i.d. centered and variance 1 entries that satisfy uniformly a sub-exponential tail inequality are included in this class, yielding results for m-sparse vectors with $m \leq Cn/\log^2(cN/n)$. Study of matrices with i.i.d. log-concave columns leads to a result on a very "high" neighborliness of centrally symmetric polytopes generated by N i.i.d. random vectors uniformly distributed on a convex body $K \subset \mathbb{R}^n$. Another interesting case is when independent vectors $X_1, \ldots, X_N$ lie on a common sphere. This is joint work with R. Adamczak, A. Pajor, and N. Tomczak-Jaegermann.

Speaker: **Po-Ru Loh** (MIT)
Title: *A numerical linear algebra view of the Tao-Vu smallest singular value limit and the SDO extension*
Abstract: We revisit the limiting distribution of the smallest singular value σ_n of an $n \times n$ random matrix with iid mean zero, variance 1 entries. Edelman performed extensive computations leaving little doubt that as n approaches infinity, $\sqrt{n}\sigma_n$ approaches a limiting distribution independent of the distribution from which the entries are drawn. Twenty years later, Tao and Vu proved this formally (under a finite moment assumption) using an approach motivated by "property testing" from theoretical computer science. We will dissect the result from a numerical analyst's perspective, starting with a reformulation of the intuition in terms of sampling a submatrix from the QR decomposition. We then explore data comparing the strengths of bounds and accuracy of estimators to reality – which we hope will pique the interest of the experimentally minded – and also argue a link (via bidiagonalization) to an underlying stochastic operator – an alternative approach which should interest the theoretically oriented. This is joint work with Alan Edelman.

Speaker: **Michael W. Mahoney** (Stanford University)
Title: *Learning on large informatics graphs*
Abstract: Recent empirical results on the structural properties of large social and information networks demonstrate that these networks are particularly ill-suited for analysis with many traditional machine learning and data analysis tools. At root, this has to do with the fact that the relationship between structures that may be interpreted as geometric and structures that exhibit empirical signatures of quasi-randomness is substantially more complex in large social and information networks than it is in many more traditional classes of data. After briefly summarizing these empirical results, as well as how they were obtained, I will describe some of the implications of these results for the algorithmic and statistical modeling of large informatics graphs, as well as for performing machine learning and inference on these graphs.

Speaker: **Per-Gunnar Martinsson** (University of Colorado, Boulder)
Title: *Enabling very large-scale computations via randomization*

Abstract: Very large data sets arise in medical imaging, in analyzing large networks such as the World Wide Web, in image and video processing, and in many other applications. Existing deterministic algorithms for extracting information from such data sets are very accurate and robust, but are often designed for a single processor computer with the data available in fast Random Access Memory (RAM). Future software must be able to fully exploit modern hardware dominated by multiple processors, capacious but slow memory, streaming data, etc. (Not coincidentally, such hardware is exceptionally well suited for generating huge data sets.)

The talk will describe a set of recently developed techniques for standard linear algebraic computations (such as computing a partial singular value decomposition of a matrix) that are very well suited for implementation on multi-core or other parallel architectures and for processing data stored outside of RAM, or streamed. These techniques use randomized sampling to reduce the effective dimensionality of the data. Remarkably, randomized sampling does not only loosen communication constraints, but does so while maintaining, or even improving, the accuracy and robustness of existing deterministic techniques.

Speaker: **Brendan D. McKay** (Australian National University)
Title: *Combinatorial estimates by the switching method*
Abstract: The method of switchings is a standard tool for enumerative and probabilistic applications in combinatorics. In its simplest form, it analyses a relation between two sets to estimate the ratio of their sizes. In a more complicated setting, there are a family of sets connected by some relations. By bounding properties of the relations, bounds can be inferred on the relative sizes of the sets. When the digraph of sets and relations is a path, precise asymptotic counting is often possible, and several of the best results for sparse enumeration have been obtained in this manner. For more complicated digraphs, useful tail bounds can still often be obtained. In this talk we extend the treatment of Fack and McKay (2007) to allow an arbitrary digraph. The main advantage is the very wide applicability, as we shall show with diverse examples.

Speaker: **James Mingo** (Queen's University)
Title: *Sharp Bounds for Sums Associated to Graphs of Matrices*
Abstract: Let $T_1, \dots T_m$ be $N \times N$ matrices and π be a partition of $[2m]$. Let

$$S_\pi(N) = \sum_{j_1,\dots,j_{2m}=1}^{N} t^{(1)}_{j_1 j_2} \cdots t^{(m)}_{j_{2m-1} j_{2m}}$$

where the indices $j_1, \dots j_{2m}$ in the sum are constrained to satisfy $j_r = j_s$ whenever r and s are in the same block of π. Let $r(\pi)$ be the smallest integer such that for all N and for all $T_1, \dots, T_m$

$$|S_\pi(N)| \le N^{r(\pi)} \|T_1\| \cdots \|T_m\|$$

where $\|T\|$ is the operator norm of the matrix T. We show that $r(\pi)$ can be easily computed from a graph associated to π. Sums of the form $S_\pi(N)$ arise naturally in the consideration of the asymptotic properties of products of random and deterministic matrices. This is joint work with Roland Speicher.

Speaker: **Raj Rao Nadakuditi** (University of Michigan)
Title: *Eigenvalue and eigenvector phase transitions due to perturbation or randomization*
Abstract: Motivated by applications in statistical signal processing and randomized numerical linear algebra, we consider the eigenvalues and eigenvectors of finite, low rank perturbations of random matrices. We uncover a remarkable phase transition phenomenon whereby the large matrix limit of the extreme eigenvalues of the perturbed matrix differs from that of the original matrix if and only if the eigenvalues of the perturbing matrix are above a certain critical threshold. It turns out that the critical threshold can be computed explicitly in closed form and is intimately related to integral transforms that arise in free probability theory. This is joint work with Florent Benaych-Georges.

Speaker: **Jonathan Novak** (Queen's University)
Title: *Restricted permutation matrices and Mehta's integral*
Abstract: Mehta's integral is a certain multidimensional integral of the type A Weyl chamber. It depends on the positive parameter β, and can be thought of as the partition function for Dyson's Coulomb gas on the real line at (inverse) temperature β. In the 1960's, Bombieri showed how to approximate Meht'as integral by the large N limit of Selberg integrals, and thus evaluate it for all beta using Selberg's Integral Formula. Later in the 1980's, Regev observed that Mehta's integral may alternatively be approximated by the large N limit of certain combinatorial sums. However, it is not known how to obtain the asymptotics of these sums directly, so this observation has not been used to evaluate Mehta's integral. We will show how to directly obtain the asymptotics of Regev's sum for $\beta = 2$ (where, by the RSK correspondence, it enumerates permutation matrices under a constraint), and thus get a combinatorial evaluation of Mehta's integral at $\beta = 2$. This is based on a certain elementary symmetry in the energy of the Coulomb gas

Speaker: **Patrick Perry** (Harvard University)
Title: *Cross-Validation for Principal Components Analysis*
Abstract: Cross-validation (CV) is a popular method for model-selection. Unfortunately, it is not immediately obvious how to apply CV to unsupervised or exploratory contexts. This talk discusses some extensions of cross-validation to unsupervised learning, specifically focusing on the problem of choosing how many principal components to keep. We introduce the latent factor model, define an objective criterion, and show how CV can be used to estimate the intrinsic dimensionality of a data set. Through both simulation and theory, we demonstrate that cross-validation can be valuable tool for unsupervised learning.

Speaker: **Yaniv Plan** (California Institute of Technology)
Title: *Tight Oracle Bounds for Low-rank Matrix Recovery from a Minimal Number of Random Measurements*
Abstract: This talk discusses several novel theoretical results regarding the recovery of a low-rank matrix from just a few measurements consisting of linear combinations of the matrix entries.We show that properly constrained nuclear-norm minimization stably recovers a low-rank matrix from a constant number of noisy measurements per degree of freedom. Further, the recovery error from noisy data is within a constant of three targets: 1) the minimax risk, 2) an oracle error that would be available if the column space of the matrix were known, and 3) a more adaptive oracle error which would be available with the knowledge of the column space corresponding to the part of the matrix that stands above the noise. Lastly, the error bounds regarding low-rank matrices are extended to provide an error bound when the matrix has full rank with decaying singular values. The analysisis based on the restricted isometry property (RIP).

Speaker: **Victor Preciado** (University of Pennsylvania)
Title: *How Relevant are Local Measurement in Complex Network Modeling?*
Abstract: An open debate in the field of Network Science is the relevance of certain 'local' measurements, such as the degree distribution or the clustering coefficients, in the modeling of large-scale complex networks. In our work, we propose a systematic approach to quantitatively determine the relevance of these local measurements. In particular, we study the effect of the most popular measurements on the spectral moments of the Laplacian matrix of the network. Furthermore, our approach provides us with a complete characterization of what measurements are most relevant to characterize the Laplacian spectrum.

Speaker: **Brian Rider** (University of Colorado, Boulder)
Title: *Small deviations for beta ensembles*
Abstract: We'll show how the (sparse) tridiagonal models of Dumitriu-Edelman may be used to prove small deviation estimates for the extremal (soft edge) eigenvalues of the general beta ensembles of random matrix theory. This is joint work with Michel Ledoux (Toulouse III).

Speaker: **Tim Rogers** (King's College, London)
Title: *Sparse and Non-Hermitian Random Matrix Theory*
Abstract: Both the introduction of sparseness and the breaking of Hermiticity are well known to enormously complicate the analysis of random matrices, as many of the best tools of the theory break down in these circumstances. In this talk I will give an introduction to a pair of techniques which can be used to handle these problems; the cavity approach to sparse interacting systems and the quaternionic Green's function method for non-Hermitian random matrices. As a motivating example, these techniques are combined to derive a non-Hermitian analogue of McKay's law for the spectral density of directed random regular graphs.

Speaker: **Mark Rudelson** (University of Missouri)
Title: *Random conjunctions matrices*
Abstract: Consider an $d \times n$ matrix A, whose entries are independent 0, 1 random variables. For a given number k we construct a $\binom{d}{k}$ times n random conjunction matrix B, whose rows are entry-wise products of k rows of A. Such matrices appear in various instances in privacy protection problems. In particular, they are used to analyze contingency tables, which are a method of releasing statistical summaries of categorical data. In these questions the parameter of interest is the condition number of a matrix, which characterizes the distortion of the Euclidean norm under the action of the matrix. We show that if the number of rows of a random conjunction matrix is significantly bigger than the number of columns, then the condition number of it behaves like that for a random matrix with independent entries. This bound is used to estimate the amount of noise necessary to protect sensitive data, while releasing statistical summaries of it. Joint work with Shiva Kasiviswanathan, Adam Smith, and Jon Ullman

Speaker: **Arnab Sen** (University of California, Berkeley)
Title: *Spectra of Large Random Trees*
Abstract: We consider the spectral distribution of the adjacency matrix for a wide variety of random trees which include, for example, preferential attachment trees, random recursive trees, random binary trees, uniform random trees etc. Using soft arguments, we show that the empirical spectral distribution for a number of different random tree models converges to a non-random (model dependent) distribution. Though it is hard to identify the limiting distributions in general, we have been able to settle some of the questions which arise naturally from the simulations. For example, for the most of the random tree models we consider, the limiting spectral distribution has a set of atoms that is dense in the real line. We obtain precise asymptotics on the mass assigned to zero by the empirical spectral measures via the connection with the cardinality of a maximum matching. For the linear preferential attachment model with parameter $a > -1$, we show that the suitably rescaled k largest eigenvalues converge jointly.Joint work with Shankar Bhamidi and Steve Evans.

Speaker: **Stephen A. Vavasis** (University of Waterloo)
Title: *Convex relaxation for the clique, biclique and clustering problem*
Abstract: We consider the clique, biclique, and clustering problems in the case that the problem instance consists of a clique, biclique, or perfectly clustered data plus some noisy data. The noisy data may be inserted either by an adversary or at random. We show that instances constructed in this manner may be solved by convex relaxation even though clique, biclique, and clustering are all NP-hard. In the case of clique and biclique, our convex relaxation uses the nuclear norm, which has recently been proved in a series of papers to exactly solve the NP-hard matrix completion problem for instances that are constructed in a similar manner. This talk represents joint work with B. Ames of University of Waterloo.

Speaker: **Balint Virag** (University of Toronto)
Title: *Spiked matrices, the stochastic Airy operator and the Painleve equation*
Abstract: The structure of some large data sets with a single trend is modeled by spiked real Wishart matrices. In the large-size limit, the top eigenvalue exhibits a phase transition as a function of the strength

of the trend. We show that the distribution of the top eigenvalue near the phase transition has a limit and express it in terms of the stochastic Airy operator. In the complex case, our results lead to a simple derivation of the Painleve equation for the Tracy-Widom distribution. This is join work with Alex Bloemendal (Toronto).

Speaker: **Rachel Ward** (Courant Institute of Mathematical Sciences)
Title: *Sparse Legendre expansions via l1 minimization* Abstract: We extend compressive sensing results concerning the recovery of sparse trigonometric polynomials from few point samples to the recovery of polynomials having a sparse expansion in Legendre basis. In particular, we show that recovery of a polynomial whose coefficients are s-sparse in Legendre basis is guaranteed with high probability, using a number of samples that scales linearly with the sparsity level. This is joint work with Holger Rauhut.